Lecture - 32
Next-Generation Sequencing Technology- Ion Torrent™

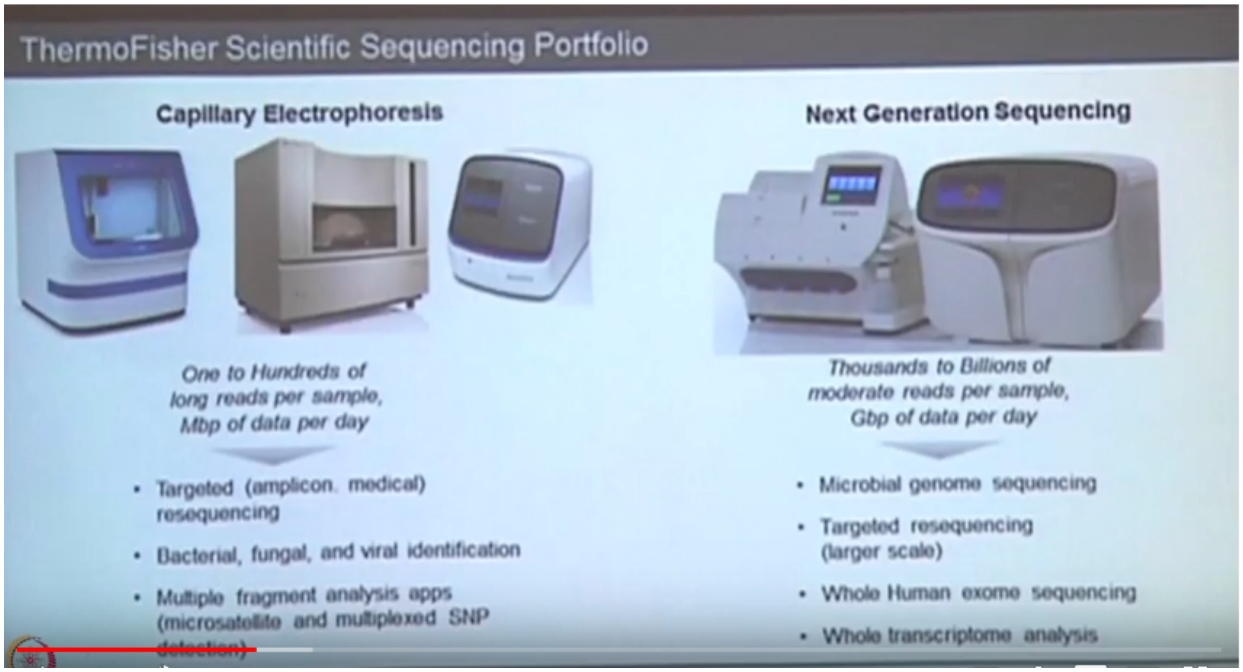
Welcome to MOOC codes, on applications of, Interactomics, using Genomics and Proteomics technologies. Today we are going to talk about, next generation sequencing technologies, before we talk about, this you know revolutionary technology platform, get to share a story with you. So, when I was a grad student like you, I was finishing my masters, in year around 2000, 2001 that time the, various genome sequence projects were, getting accomplished, it was in a big new that time, big breakthrough

that time that the human genome project, the first draft, was completed. And many other model organism and their sequences, were also getting completed at the same time. Just imagine that you know first time, we were getting information that what you know, makes it human what our genetic composition is about, it was you know really big, science breakthrough, because first time, we now know that how many genes are, actually available to us, which makes it a human, which makes us so, unique. And different than other model organisms, before that they were only speculation that you know what makes human and you know how many genes you may have, but the first time, the evidence for genes and the virginity composition, the genome, was first time revealed. Right? It was differently big revolution, in the science field, in general and definitely in the biology and medicine area, but it's not ended. Right? It is not finished that time so, many times a technology, showed us a promise and then you know it fades away, but what has happened in this area of genomics, that you know 2002 and 3, first trying the draft human genome were released, but since then, there has been tremendous advancement which has happened, in this field of genomics, reaching to where we are which we say as a next-generation sequencing technology.

The series of advancement happened for the first generation, second generation, third generation of resolution sequencing technology. And I must elevate that advancement, made by engineers and technologists big data scientist and biologists together, has really brought, where we are currently, in our magnitude of doing sequencing based experiments. So, what has changed from 2002 in 2018 main things has changed, is our speed, the speed at which, we can now sequence the genome, imagine the human genome projects, took more than 10 years time to accomplish. Now within two days time or 48 hours time, you can do sequencing of a given, individual with the next generation sequencing technology. That humongous change, additionally what $2,000,000,000 took accomplished one human, genome sequence? Now it takes maybe only $1,000, to accomplish the same sequencing technology based work. So, both cost wise, as well as the speed wise, as well as accuracy wise, the net generation sequencing, moving from the Sanger sequencing, based platform, have made tremendous revolution. So, I'm sure now you're excited, to know that what NGS technologies are, which are the currently, big industry player, big industry technologies, which are leading this field and in this light. We have invited a guest Dr. Atima Agarwal from thermo Fisher scientist, who is manager a commercial service, training and tech support at Thermo Fisher Scientific, who will be taking this lecture? And she will try to provide, you the basics of next generation sequencing, first and then give you little more detail and specifics for, one of the technologies of ion intolerant**.** So, let me welcome Dr. Atima Agarwal, for her lecture on, NGS based technologies and its applications. So, I'm Dr. Atima Agarwal from Thermo Fisher, I've been with this company since many years now. So, if you remember or if you've heard, in pre-1990 era the sequencing used to be done on gels, with radioactive labels yeah. So, and it used to be a very, very cumbersome affair, but then it had contributed a lot to the science in those terms, then first human genome was being sequenced on these systems. And there were many platforms, many of these capillary, electrophoresis platforms if you look into the pictures of Sanger Institute during that time, in lines they were these kind of systems and they were generating sequences, day and night and that's how the first human genome, got sequenced and got assembled. So, those were a lot of like I would say that, benchmarks in sequencing, suddenly somewhere during 2004 2005, we started hearing about next-generation sequencing. So, if I were to talk about, next-generation sequencing, it is more aptly called as,' Massively Parallel Sequencing'. So, with Sanger what you're doing there is that either you are running it on 96-well plates or you are running it on strip tubes. So, one well is giving you one sequence, yeah? So, one sequence, which can which is a stretch of ATGC and this can be as long as a thousand, base pairs eleven hundred base pairs and it can be as short as a hundred base pair, fragment. So, the one well, gives you one sequence.

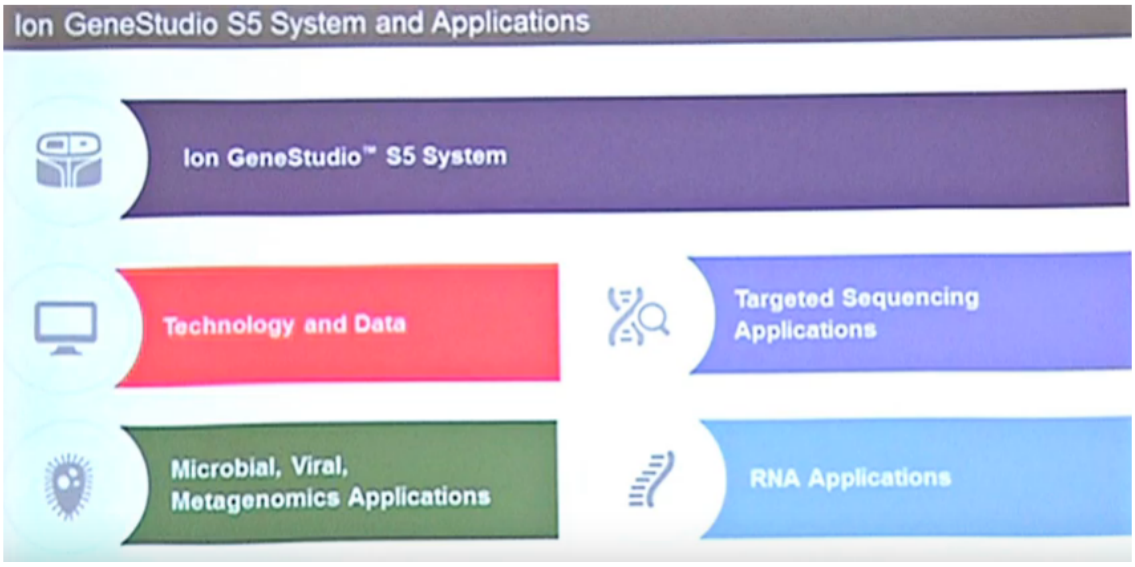
Now suddenly, with massively parallel sequencing you can imagine, what is happening is; you have millions of wells and you have this sequencing per say, happening in all those millions of fills and that is how you are generating a lot of data per run. So, that is massively parallel sequencing.

Refer slide time :(06:31)



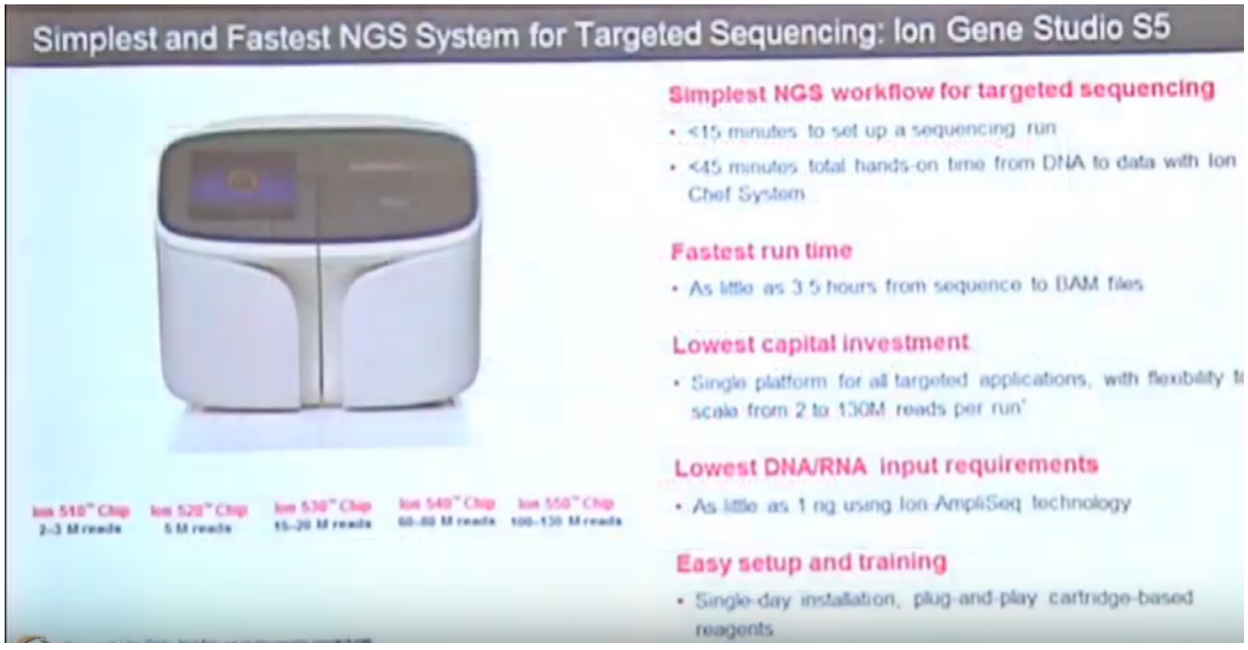
So, just to tell you that just because, we are now nowadays talking a lot about NGS, it is not that capillary sequencing, has lost its charm or has lots of lost its utility, it is still a very user friendly technique for a lot of applications. So, for tech for things, wherein you do not really require, a lot of data, things can be none simpler, for a lot of microbial research or a lot of single gene based, tests and all those things they can be still very well and very efficiently being done on capillary sequences, till they this has been the gold standard technology, for doing a lot of things, like targeted. So, you have certain Amplicons to be sequenced or you have a den over organism to be sequenced, to fill up those kind of assemblies, then you have multiple, there are multiple fragment analysis, applications wherein you are not actually, sequencing the fragment, you're just looking into this size and based on the size polymorphisms, you are telling you are giving certain scientific answers. So, these this is about capillary sequencer, coming to next-generation sequencing. So, now as I mentioned that this is a technology wherein, you have millions of wells. So, if you see that these are the chips, which are used with ion technology. And you have millions of wells on these chips. So, millions of wells so, you can imagine that each well has a sequence and that is how you are generating millions of sequences in a run.

Refer slide time :(08:14)



So, the agenda, how we've planned out the session for you is that we are going to talk about, our new series of NGS platforms, which is ion gene studio is five systems. The technology behind the gene studio s5 and some of the applications. So, this is the fastest sequencer, which is available, in terms of an NGS, which can produce starting from once you've started your sequencing run, it can produce data in as little as 3.5 hours, from a sequencing run.

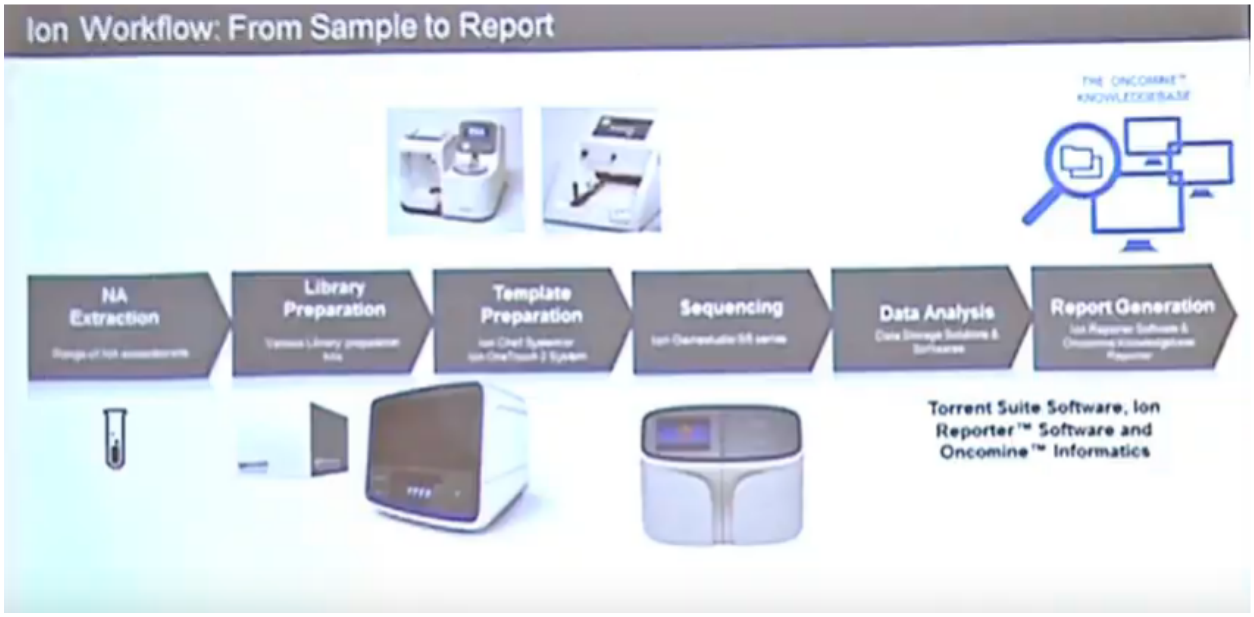
Refer slide time :(08:48)



There are lots of different kinds of applications people want to work on NGS. And a lot of these samples are very challenging samples, like when you extract DNA or RNA from FFP blocks, when you are

working with liquid biopsies, like you do not you really are looking at cell free DNA yeah. Or you have very little RNA, from a cell and still you want to study the whole Transcriptome of the cell. So, our technology, is such that it is compatible, with as low as 1 nano gram of DNA or RNA and the backbone of this technology, is Amph seq technology, will come to those details later.

Refer slide time :(09:27)



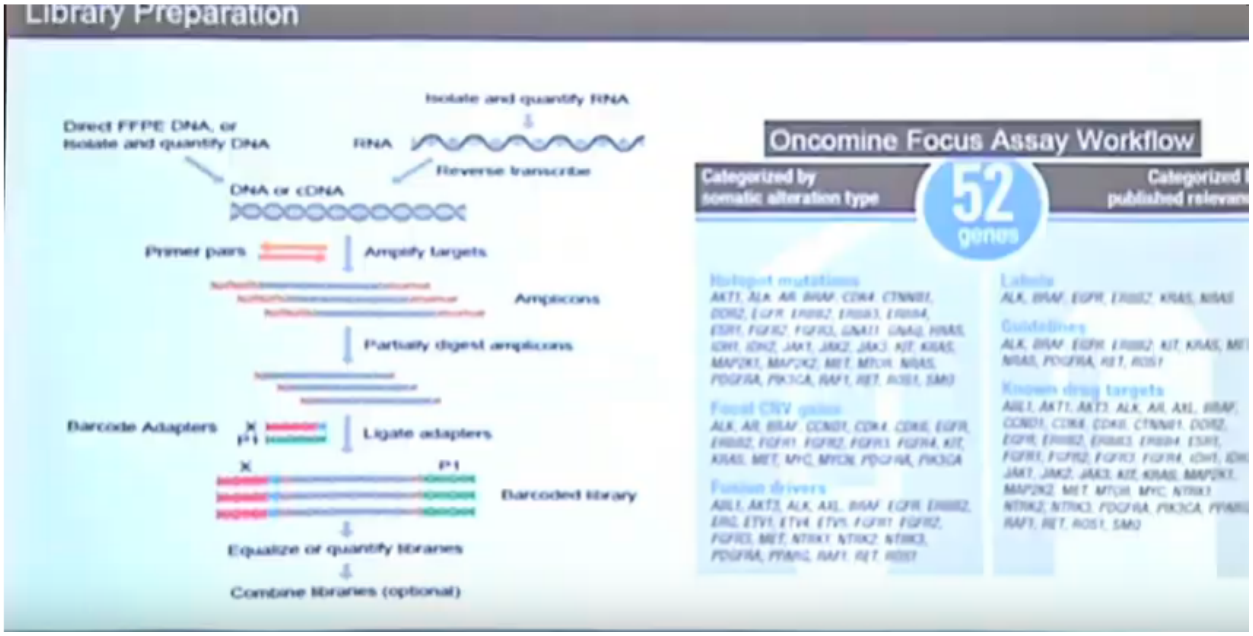
So, coming to the workflow now. So, how we are generating these millions of reads. So, you start it nucleic acid, now this nucleic acid can be DNA, can be RNA. So, where do you think you, would be starting with the DNA, what kind of experiments do you think you would be starting with the DNA, on NGS, DNA as a sample, whole genome sequencing, yes plasmids yes, targeting read sequencing. So, when you are not so targeted read sequencing is basically when you do not want to sequence the whole genomes, you just want to sequence some targeted regions of the genome, that is targeted read sequencing, yeah. So, where do you think you would be using RNA as your starting material Transcriptomes yeah, again like targeted, targeted transit, you are looking at some targeted transcripts, are you looking at some targeted fusion events, yeah, anything else, yeah RNA seq and RNA seq typically why would you do an RNA seq, you want a sequence, yeah. So, this DNA is transcribed into RNA and then hence there are more probabilities that this particular part of the genome, is actually getting transcribed on, further translated into a protein, but that's just a possibility, there are many long there many non-coding RNAs also yeah, it's like small RNA yeah, yeah pictures yeah. So, which are they're involved in, lot of regulation activities, of how the Transcriptome is shaping, within a cell yeah. So, people do a lot of different kinds of experiments, now for your what you would look out, for in your NGS is that all these experiments, because the data requirement for all these experiments, is very different yeah. So, when you are doing a seek genome, it's a very straightforward calculation that you generally tend to, generate 100x data. So, whatever is the genome size, you multiply by 100 there are definitely, ifs and buts around, it will not go into those details but, then the data requirements for these different kinds of experiments are very different, yeah. So, that is where you need that flexibility, that single NGS system, can cater to all those requirements. So, you started with your DNA or RNA yeah to begin with.

So, you isolate a DNA RNA from cells or from some culture isolates or whatever. So, now once you've isolated this DNA and RNA the next step is that you prepare a library. Now library is basically a collection of fragments, whether it is coming from DNA or coming from RNA, which you finally, want to sequence, yeah. So, this can be an enriched portion, yeah. So, for targeted sequencing, you would need to enrich that those targets first and then only you can go for the sequencing, for whole genome sequencing, since your aim is to sequence whole genomes. So, you do not need any enrichment, you will just extract DNA and you will go forward for library preparation. So, library preparation, is basically collection of fragments, whether DNA or RNA, which you finally want to sequence, yeah. So, what, what the basic aim in library preparation is that, you are now these next-generation sequences, you we talked about that you are generating 200 base pair it reads, you are generating 400 base pair reads you are generating 600 base pair reads. So, supposingly, we take an example of whole genome sequencing. Now what you're going to do is with whole genome sequencing, you want your aim is to sequence that genome as fast as possible yeah and with as bigger reads as possible. So, probably you will pick up a 400 or 600 base pair chemistry and so, what you need to do with the genome is, you shear this genome yeah you break this genome into smaller pieces. So, that can be done by, enzymatic treatment or that can be done by sonication. So, once you've broken down this genome. Now these are the fragments which you finally want a sequence, but you don't know what is there. So, you need to have some adapters, yeah or linkers, which will help you facilitate, sequencing these regions, yeah. So, in library preparation, what you're doing is? You are enriching? The region of interest and second is you are like getting it, at both the ends you are like getting it with linkle's these linkers are double-stranded DNA fragments, which are basically exploited. So, that the primer can then sit on these kind of fragments and then. So, you have your fragment of interest in between, you link the adapters at both the ends and so, why?

So, that leave now the these adapter sequences, you know, because these are coming from a chemistry and we tell you that these are the adapter sequences. And then the primers comes in sits on these adapters. So, that you can sequence, this region in between, yeah. So, that is basically so, with library preparation, what you are doing is? You are enriching? Your region of interest and you are like getting these fragments to adapters yeah. Then comes template preparation. So, basically, now these are the things which comes, which are specific to a technology. So, in template preparation, what you're doing is? That you taking these library molecules and you are amplifying onto ion spheres. So, we are talking about, ion technology. So, we're using ions spheres, as the mode, these ions spheres, get coated onto those or get loaded, onto those chips and finally these ions spheres are carrying those regions of interest or those library molecules, from which you are generating all those reads ,yeah. So, that process of taking, these library molecules and amplifying them onto the ion sphere, is called,' Template Preparation', then comes sequencing. Now you have your ion spheres, which are template, which have library molecules, amplified onto them and now you want a sequence. So, you load these ions sphere, onto the chip and then you generate the data. So, once you've loaded, you initialize your sequencer, which we discussed this gonna take only 15 minutes. And then your sequencing run goes for anywhere from 2.5 to 4.5 hours, depending on the kind of the read length you're targeting. So, talking about the, instrumentation, which is required in this workflow. So, we have it's the library preparation, can be done manually, can be done on ion Chef, there are some kind some parts of library, which can be done on Ion Chef, yeah. So, this is an automated way of preparing libraries, otherwise you have manual way, of preparing libraries, wherein you are using certain magnetic beats, certain library preparation, kits from us and the process generally, for targeted sequencing or for whole genome sequencing, the process is around four to four and a half hours for manual and probably for RNA seq, it's a days protocol, because you have to do certain levels of enrichment and you

and you, have to do more QCs in that. So, if you talk about manual preparing libraries manually. So, that's the kind of time stream, otherwise you can use Ion Chef, for preparing your targeted sequencing libraries, automatedly then comes the template preparation. Now as we discussed, during template preparation, you are loading these library molecules onto the ions ion spheres and then finally these ions phase are loaded onto the chip. Now these loaded chips, you take from here and you just load it in on this system and they are good to go for sequencing.

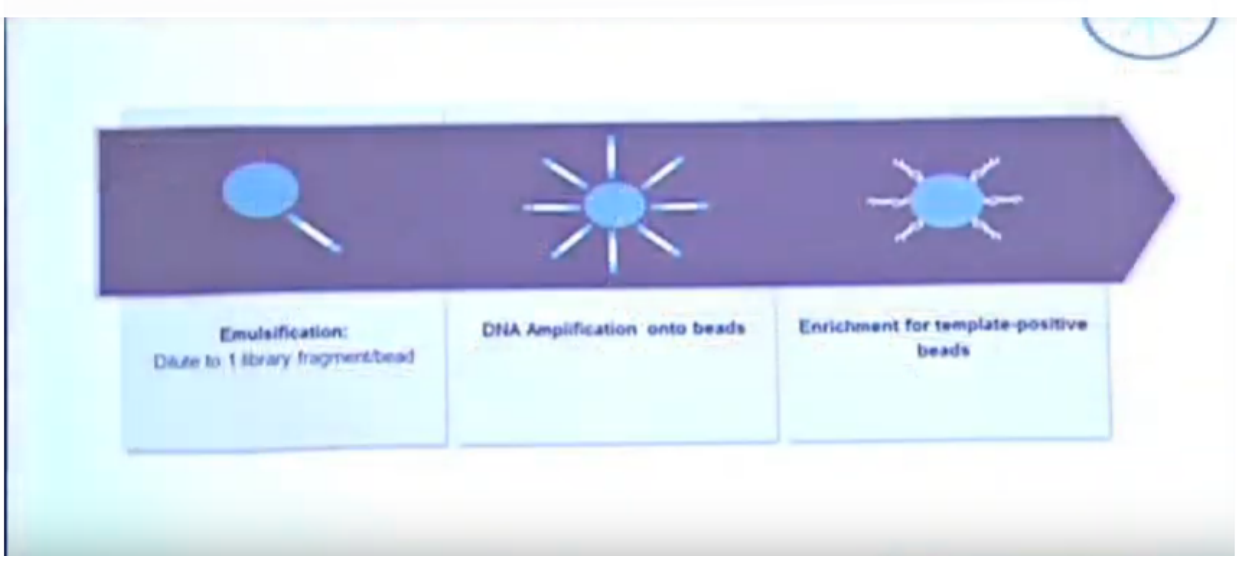
Refer slide time :(17:57)



So, just coming into few details on, library preparation. Now how we go about library preparation is, since you are targeting, variations at the DNA level, as well as at the RNA level you isolate DNA and RNA yeah. Then you reverse transcribe, this RNA. So, now you have DNA or cDNA. So, you take this primer pool and you use this DNA and cDNA as your target, all these genes, yeah get amplified at one go and then since the primers are coming from conserved, regions, wherein you do not have any hotspots or any kind of variations, to be looked into we partially digest these primer sequences and then we like get adapters. Now when we say that you can do smaller experiments also. so, you have a 50 to gene panel and you have some 12 samples, to be done. So, what you're going to do is you index these 12 samples, by 12 different barcodes. Now these barcodes so, when we said that library preparation is all about, enrichment of the targets and ligating adapters. So, when we are ligating these adapters, you are using bar coded adapters, to differentiate, one sample from another, yeah. Now this bar code, is a 10 base pair unique sequence, which is present in the adapter yeah, which is the which will finally help you to differentiate, between different kinds of samples, which you are loading on one chip. So, once you have once you have, ligated these adapters you pool all these libraries. Now you are you are working with probably six samples or not so, with based on the number of samples you plan, to multiplex on a chip, you are going to use the number of barcodes, yeah. Once you've done that now, you want equal amount of data, for all

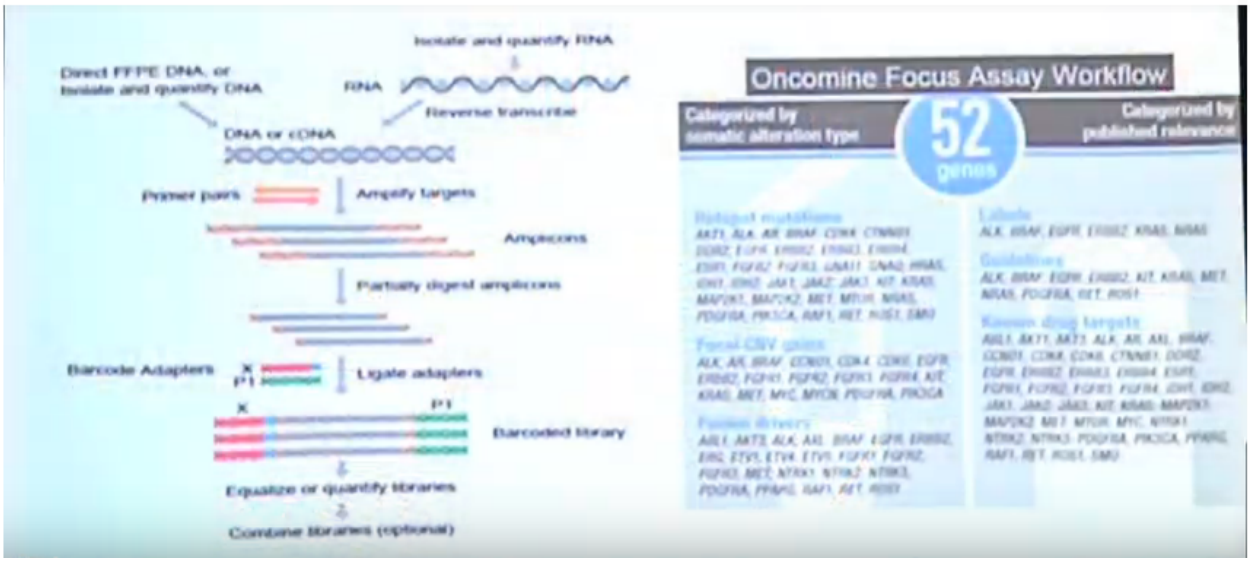
those libraries, what you're going to do is in terms of their morality concentration, you will use equal molar concentration, for all those libraries and make one pool of that, but the aim is that when you're doing small RNA sequencing, for all these samples, you are aiming that you generate equal amount of data, for all these samples but because, otherwise probably some of the samples have worked, out pretty well and some will not. So, to ensure that, you need to pool all these libraries, in aq molar proportion. So, once you've pooled that after your library preparation, step. Now you started with 12 or you started with 20 samples. Now that becomes one sample, yeah and now how are you going to differentiate, now because it is now one sample, but you started with 20 samples. So, when you generate the data, how are you going to differentiate, yes we use bar coded adapters. So, the system is first thing it's going to use, when its sequences, it is sequences the barcode it's barcode sequences and then it sequences, your region of interest and then it makes the bins that okay, this read belongs to this sample and it is in this bin.

Refer slide time :(21:16)



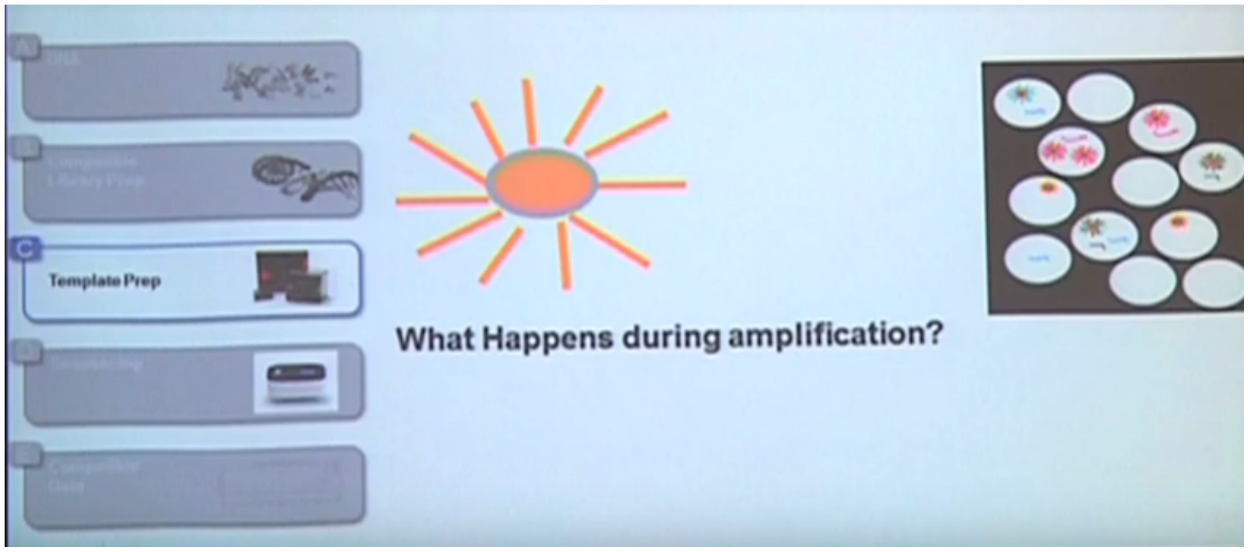
Once we are through with our library preparation, then the next step is template preparation. So, in template preparation the aim is that you clonally amplify this. So, you have prepared you started with, targeted sequencing or you started with whole genome sequencing or you started with an RNA-seq experiment, you have tons of target, millions of target molecules, which are to be sequenced. So, what we are aiming during template preparation, is that one library molecule, gets clonally amplified onto one ion sphere. So, this is the basic aim of this particular, step and so, this library molecule, gets amplified, onto this ion sphere and this is one clone, which is getting amplified, all over on the ion sphere then there could be some ion spheres, which have not got amplified yeah. So, you do not want to load them, on the chip. So, you negate all those ions spheres at the enrichment step yeah. So, what is happening is? That we are using ion spheres, which have an Oligo, which is linked to these. So, this Oligo is complementary, to one of the adaptors, which you used during library preparation. So, if we go back a few slides.

Refer slide time :(22:40)



So, we used two adapters P, one and one was bar-coded adapter. So, and we mentioned earlier, that we are basically exploiting these barcode, these adapter sequences, for us to enable the sequence, to sequence these millions of fragments later on, on one chip. So, we are using this particular p1, adapter sequence yeah, for the library molecules two complimentary go and bind onto the ion sphere. So, now if we go back to that slide.

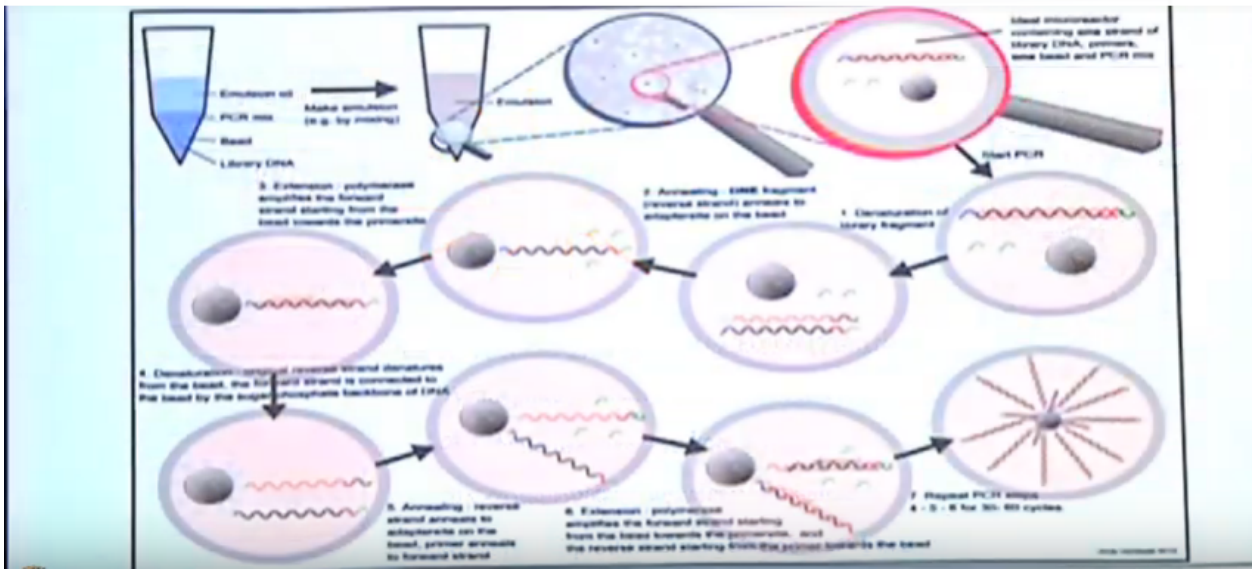
Refer slide time :(23:13)



So, these ion sphere are loaded or are coated, with all egos which are complimentary to the p1 adapter. So, when these emulsions are being created, you are basically, making these water molecules, which are surrounded by oil and you are trying to amplify, these library molecules. So, when this emulsion is being created, this is a random distribution, of these ions sphere with the library molecule and enough PCR components for it, to amplify this library molecule onto this ion sphere. So, what is happening is? A

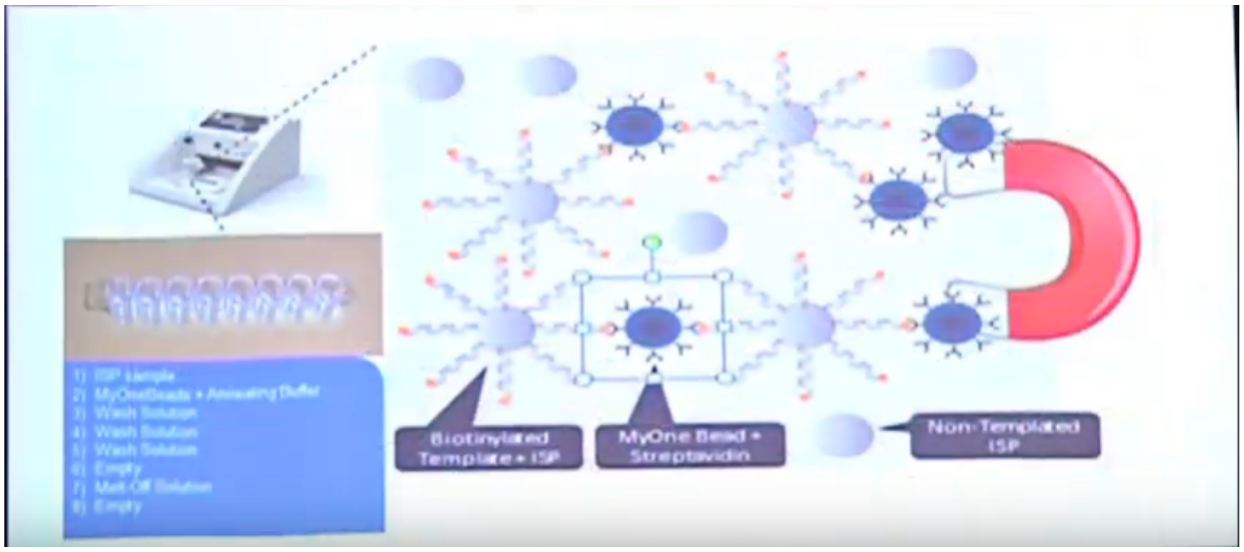
library molecule is coming, it is getting denatured and it is sitting, on this complimentary sequence, which is complimentary to the p1 adapted. And then it goes ahead and it sequences this part and that's how you sequence. Now this goes through a cycle of denaturation, annealing an extension, like your normal PCR. But when while we are talking, all this has been. So, you are not manually involved with all this process, all this is being done by these systems, either semi-automatedly or fully-automatedly. So, then when, when it gets denatured, it then again binds to another, Oligo on that same bead and you then that's how you are generating, you are clonally amplifying this particular molecule, onto the ion sphere.

Refer slide time :(23:13)



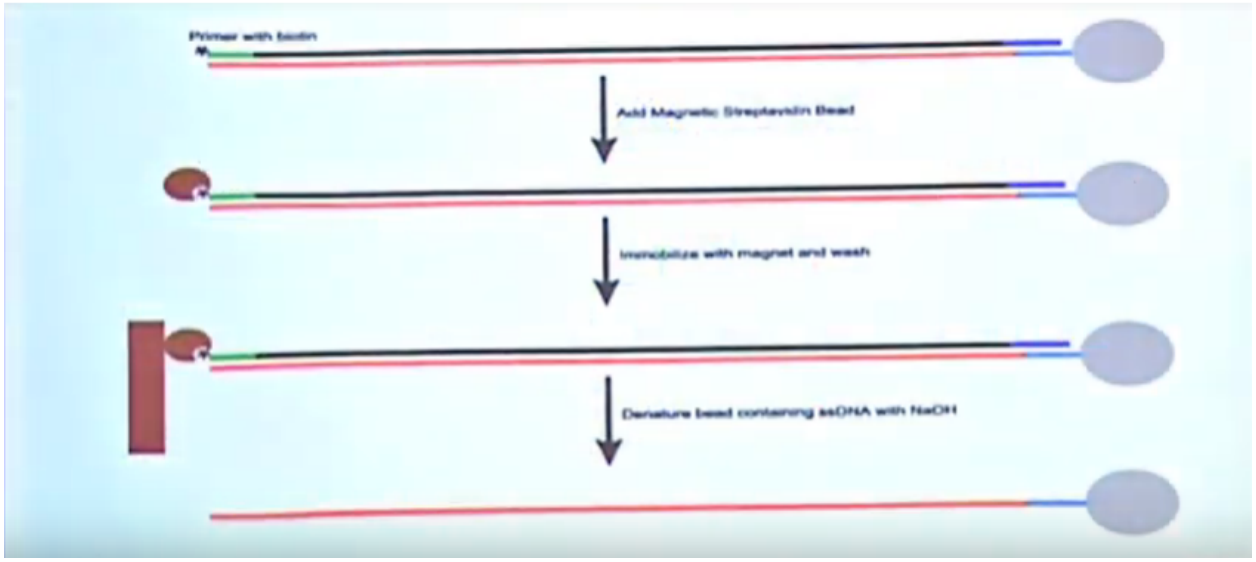
So, what is if we go into the details, what is happening is you have an emulsion oil, then you have a library molecule, you have some primers this is your ion sphere, this library molecule, gets denatured, it gets bound, to the ion sphere, it gets extended again gets denatured now this template, goes and binds to another Oligo and finally you have this one library molecule, clonally amplified on to this ion sphere yeah.

Refer slide time :(25:08)



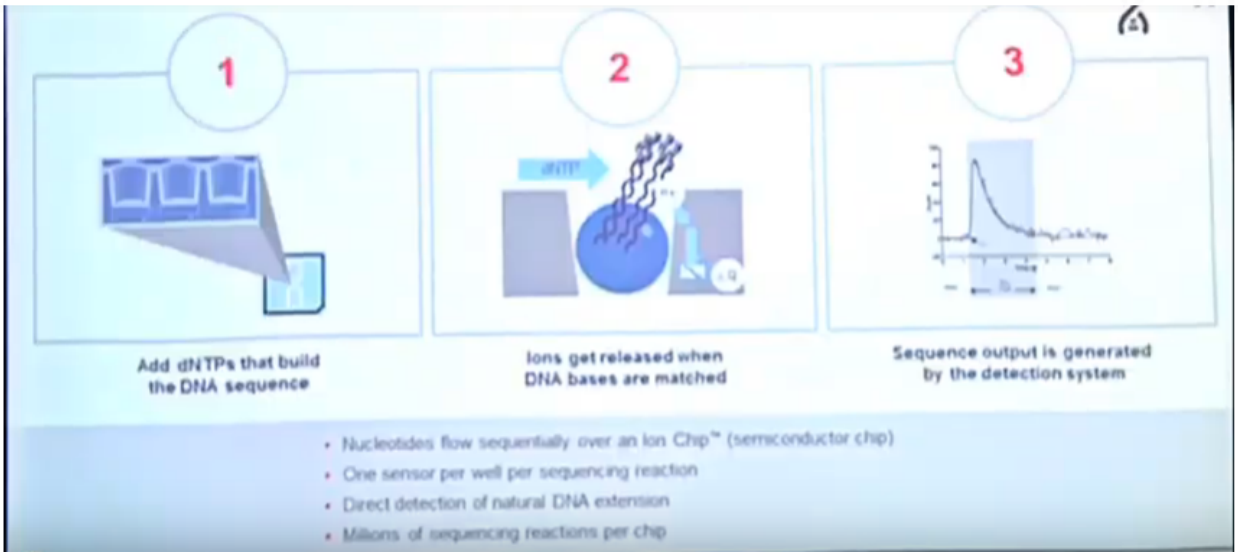
So, now we said that there could be some molecules or some of those water droplets when, the emulsion was being created they did not get a library molecule. So, they will not get amplified also, but we do not want to take them on the chip, because they are still going to occupy some space on the chip. So, why to lose on that data? So, what we are doing at the enrichment step is, when we were doing amplification, on the ions fears one of the primer was byte annihilated. So, now all the amplified ion spheres are byte annihilated, we are using Streptavidin coated beads to fish all the amplified molecules and rest all molecules remain in the solution.

Refer slide time :(25:52)



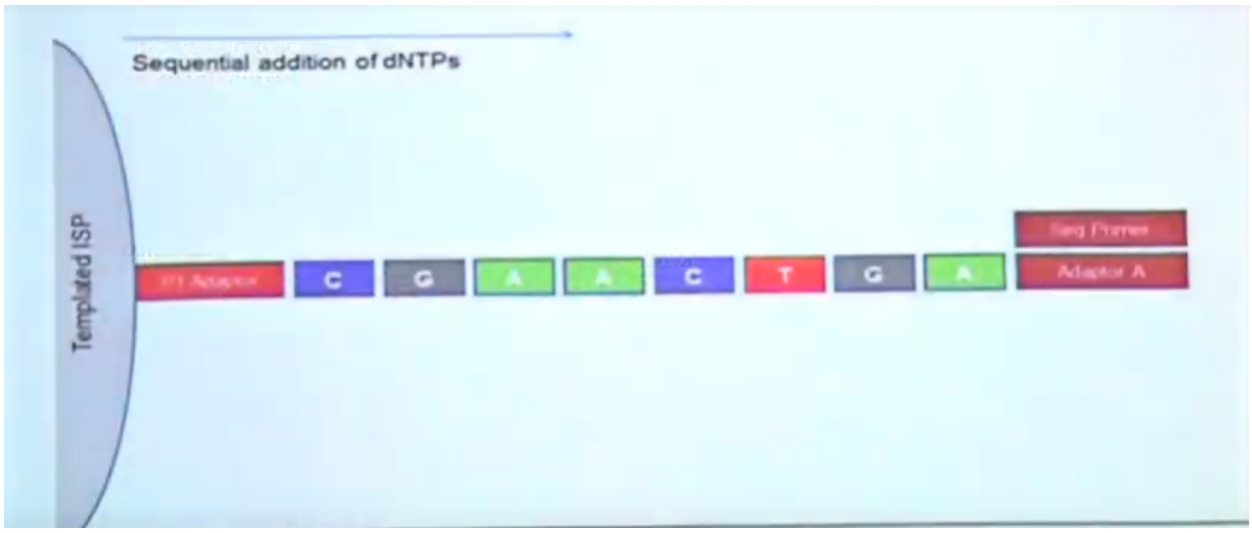
And then we just go through, so, this is primarily your ion sphere, this will have these molecules all over, because we clonally amplified. So, you use Streptavidin coated, beads and this Streptavidin cool Streptavidin coated beads are magnetic. So, with the help of a magnet all these are fished out and then, you have only the beads, which are and which have amplified, which are finally to be loaded on the chip. So, this is the enrichment part of the template preparation.

Refer slide time :(26:26)



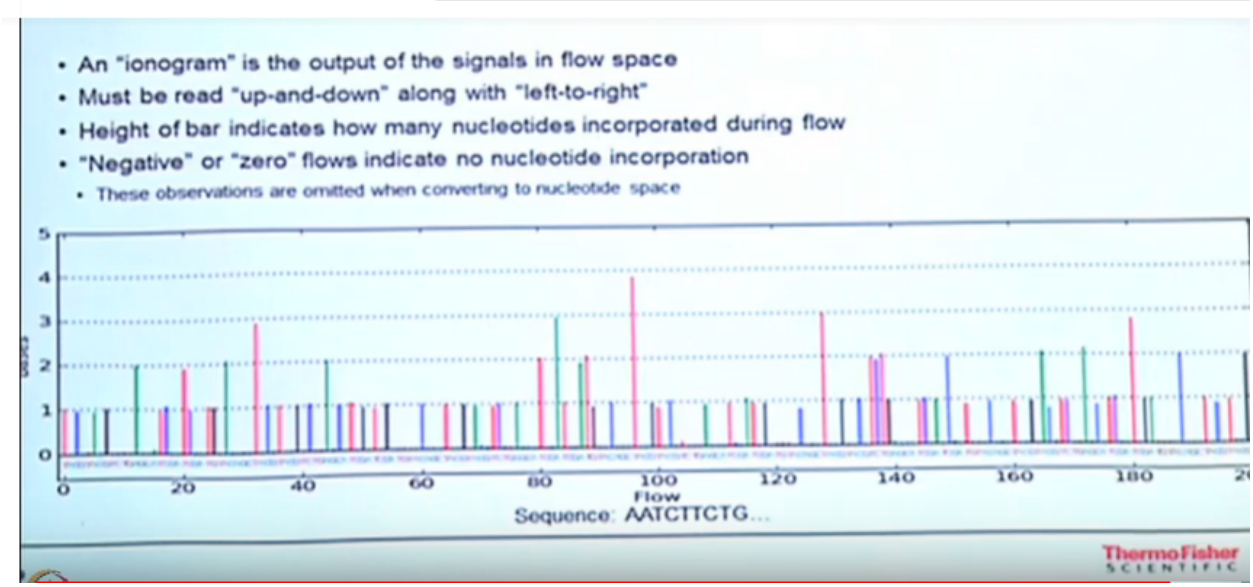
Then what is happening at the sequencing level. Now I showed you that chip that chip has millions of wells and each well, is working is generating a sequence. So, you're so, this is your chip it has millions of wells and if we are now, to consider that this is one well, this one well has one ion sphere, which is loaded with one kind of template. So, what, what is happening during sequencing is? That when we when ion chef is preparing, the sample for the sequencing, it is already adding the primer in polymerase, through the chip itself. Now what else it needs for extension, dnPTs yeah. So, what the system does is it is adding dnPTs, sequentially onto the chip. And then naturally, we know that whenever dnPT, gets incorporated or gets polymerized into a growing chain of DNA, there is a there is a bond formed, between the phosphate group and the hydroxyl group, yeah. And this Phosphodiester bond, when formed this releases a hydrogen ion yeah. So, now we coated these library molecules, onto this ion sphere. So, this same event so, supposedly the first base here, is an a yeah. So, you, you had flown DTTP. So, it will have a complement, it is a complementary nucleotide it will bind and it will release hydrogen ions and these hydrogen ions, get converted into a voltage signal, which is finally being captured. So, this ion basically is bring these release of so, many hydrogen ions and one will, are bringing up a change in pH for that particular, well yeah. And when the pH is being changed, for that particular well that change in pH is being recorded as a change, in voltage. So, basically you are converting a chemical, information into a digital information and that is how you are saying that a base a signal has been received and a base has been called. Now supposingly now, we said that, T is your or A is your first base and the system, had flown DGTP. So, now that is G is not going to bind to A yeah. So, what will happen is there, will be no binding, no hydrogen ion release. So, no, no change in Ph, no signal detected by this system.

Refer slide time :(28:59)



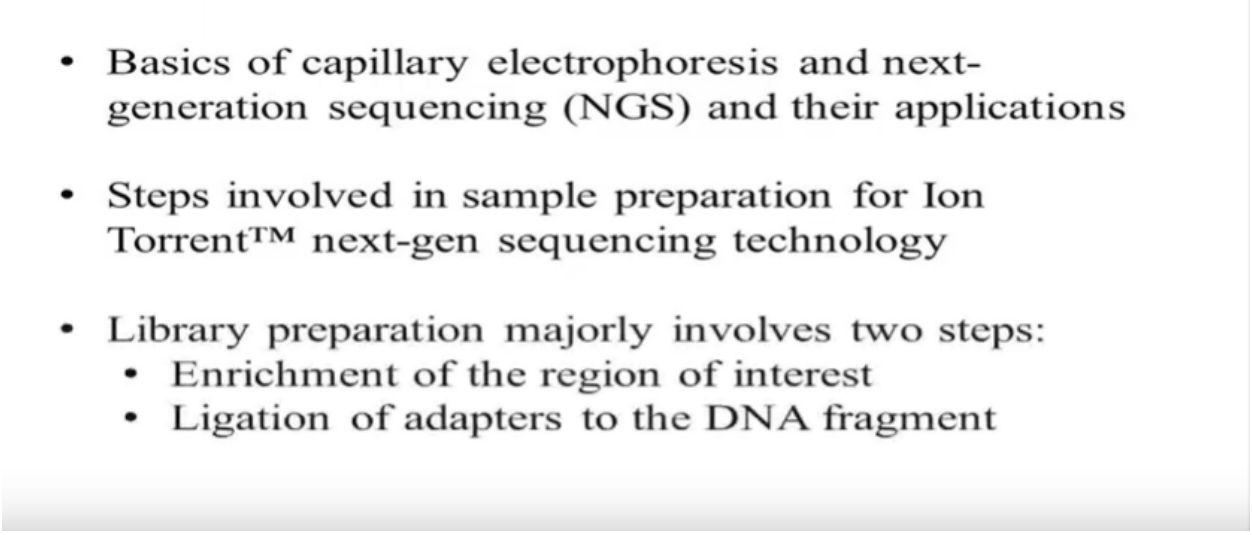
So, we will just see this by a cartoon. So, considering that this is your one ion sphere, these are your adapters yeah. This is the bar coded adapter and this is your adapter, which you used for ligating these sequences on to the ion sphere, then what is happening is your primer is coming in binding here, you add the DNC piece sequentially. Now what will happen is, this will bind; you will have a change in pH which will be recorded by the system. And the system knows that it is flowing the DTTP at this particular time and that is, why it can associate this signal with an a yeah. So, we are not using any fluorescently label dnPTs. So, this is a simple chemistry, now what will happen in the next. So, we had flown in an A and the next sequence, next target base is G, they are not complementary. So, it will not mind, it will get washed off then again. So, likewise you will build up the sequence, now what will happen at this step? Supposingly we are flowing a DTTP what will happen? Sorry yeah, since there are two ways in a stretch and there's nothing stopping it it's a growing chain, yeah. So, you will have double, the signal yeah. So, there are double molecules, double number of hydrogen ion release, which will bring up a double change, in the voltage and that is how it has been recorded as, signal, which is double and that is where you can tell that they were two consecutive, is in the sequence. So, then this by adding the disease dnPT sequentially, finally the system is being is creating a sequence, which is finally being read as an Ionogram.

Refer slide time :(31:03)



So, this Ionogram is again being read by the sequencer, this has to be read from left to right and for the number of bases. So, whether you get a signal or not that will define, the sequence of the dnPTs, the intensity of the signal will define, how many of those DNT how many of those is or T's were in a rho.

Refer slide time :(31:30)



Refer slide time :(31:37)

- Linkers/adapters are dsDNA fragments containing bar-coded regions that allows multiplexing
- Template preparation is a process that results in clonal amplification of library molecules on ion spheres
- Enrichment steps involves selection of template-positive beads with the help of biotinylated primers

Refer slide time :(31:46)

- Ion Torrent™ next-gen sequencing technology records change in pH upon addition of dNTPs
- The output signals of the sequence is known as ionogram and should be read from left to right
- The height of signal in the ionogram indicates the number of nucleotides incorporated

So, hope you got, a glimpse of next generation sequencing technologies, which has really revolutionized, the field of genomics, you were given the basics and some possible applications, using this technology platform. And when a specific technology, of Ion Torrent, was also illustrated and discussed, these concepts will be again covered in more detail, in the following lectures, when we have invited some more industry experts, representing different technology platforms, who will also provide you, the basics and the current status of these technologies. So, see you in the next lecture. Thank you.